Quantum repeaters with imperfect memories: cost and scalability

M. Razavi

Institute for Quantum Computing, and

Department of Electrical and Computer Engineering,[*]

M. Piani and N. Lütkenhaus

Institute for Quantum Computing, and

Department of Physics and Astronomy,

University of Waterloo, Waterloo, ON, Canada N2L 3G1

Abstract

Memory dephasing and its impact on the rate of entanglement generation in quantum repeaters is addressed. For systems that rely on probabilistic schemes for entanglement distribution and connection, we estimate the maximum achievable rate per employed memory for our optimized partial nesting protocol, when a large number of memories are being used in each node. The above rate scales polynomially with distance, L, if quantum memories with infinitely long coherence times are available or if we employ a fully fault-tolerant scheme. For memories with finite coherence times and no fault-tolerant protection, the above rate optimistically degrades *exponentially* in $\sqrt{L}$, regardless of the employed purification scheme. It decays, at best, exponentially in L if no purification is used.

PACS numbers: 03.67.Hk, 03.67.Bg, 42.50.Dv, 03.67.Dd

---
[*]Electronic address: mrazavi@iqc.ca

I. INTRODUCTION

Quantum repeaters enable entanglement distribution between remote parties by relying on a network of quantum memory units [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. In their seminal paper [1], Briegel et al. demonstrate how to distribute entanglement over arbitrarily long distances using ideal quantum memories as well as highly efficient quantum gates for entanglement purification [11, 12, 13]. The resources needed in such a scenario will then only grow polynomially with distance for a fixed desired fidelity for the final entangled state. Further studies have shown that so long as the coherence time τ_c of the memories is much longer than the transmission delay L/c, where L is the distance between the two parties and c is the speed of light in the channel, the above assertion still holds [7, 8]. It is not clear, however, how the required resources scale in the limit of large distances when τ_c is finite. Here, we answer this question by assuming that the only error-correction mechanism used in the system is entanglement purification in conjunction with allowing for error-free, but probabilistic, gates to be used instead of erroneous deterministic ones. We find that, even under optimistic assumptions, the system cost explodes as a power of $\exp[\sqrt{(L/c)/\tau_c}]$ for $\tau_c \ll L/c$, unless we employ a fault-tolerant scheme to remedy the memory decay [7, 18].

In order to quantitatively address the cost factor in quantum repeaters, we look at the generation rate of maximally entangled states *per employed memory* in the system. We employ proper entanglement measures [14], instead of merely looking at the fidelity, to find this rate. We obtain this rate, in the steady state, assuming that the resources in our system are being successively used, according to a proper protocol, to create entangled states. Such a rate-over-cost measure is useful for applications in quantum key distribution (QKD) [15], where the generation rate of secure key bits is proportional to the rate of entanglement generation. Moreover, it provides us with a fair and practical measure for comparing different quantum repeater setups and their contrast with alternative schemes for entanglement distribution that do not rely on using quantum memories, such as quantum relay structures [16] or the direct transmission of entangled photons. In the latter cases, the rate will decay exponentially with distance as a result of loss in the channel.

Memory decay is one of the most challenging problems in quantum repeater technology [17]. Its deteriorating effect, however, has not yet fully scrutinized. In [7], authors study the role of memory errors in quantum repeaters, but they treat the required initial entanglement

as a given resource. This approach cannot fully capture the memory decay problem as it neglects to account for the corresponding waiting times during initial entanglement distribution. Entanglement distribution, regardless of the employed scheme, is a probabilistic process, mainly because of its dealing with the loss in the channel, and therefore, the time required to entangle two memories is a random variable. Collins et al. [8] consider this probabilistic nature in a multiple-memory configuration, and report a numerical-analytical rate analysis that includes the effects of memory decay. They model the memory decay by associating a lifetime window—within which the stored entanglement is unaffected and beyond which it is destroyed—to each memory. This simple model is not, however, sufficiently realistic to properly account for the effect of memory errors on the rate, especially in the regime of short coherence times.

In this paper, we introduce a partial nesting protocol, inspired by the blind repeater protocol proposed in [7], and find the throughput in the two cases of perfect and imperfect memories, with or without purification, for a generic quantum repeater system that relies on probabilistic schemes for initial entanglement distribution as well as for its entanglement connection. Such probabilistic architectures for quantum repeaters [2, 5, 6] can adapt themselves to post-selection-based self-purification mechanisms, which—in the absence of memory errors—will allow them to create high-fidelity entangled states over moderately long distances without relying on high-quality quantum gates [3]. Such an advantage comes at the price of their achieving relatively low entanglement generation rates owing to the probabilistic nature of their operation. They are nevertheless attractive options for QKD systems whose security over long distances can be guaranteed by a combination of such quantum repeater links and sparsely located trusted nodes. To account for the above practical issue, in Sec. II, we describe the probabilistic framework of our multi-memory quantum repeater system. In Sec. III, we present our rate analysis for different scenarios. Numerical results will be presented in Sec. IV, and we conclude the paper in Sec. V.

II.   SYSTEM DESCRIPTION

In our analysis, we study a quantum repeater link of length L with 2^n sublinks of length L_0, as shown schematically in Fig. 1, with n being the nesting level of the system. At the end of each sublink there is a bank of N quantum memories. We schematically refer to the
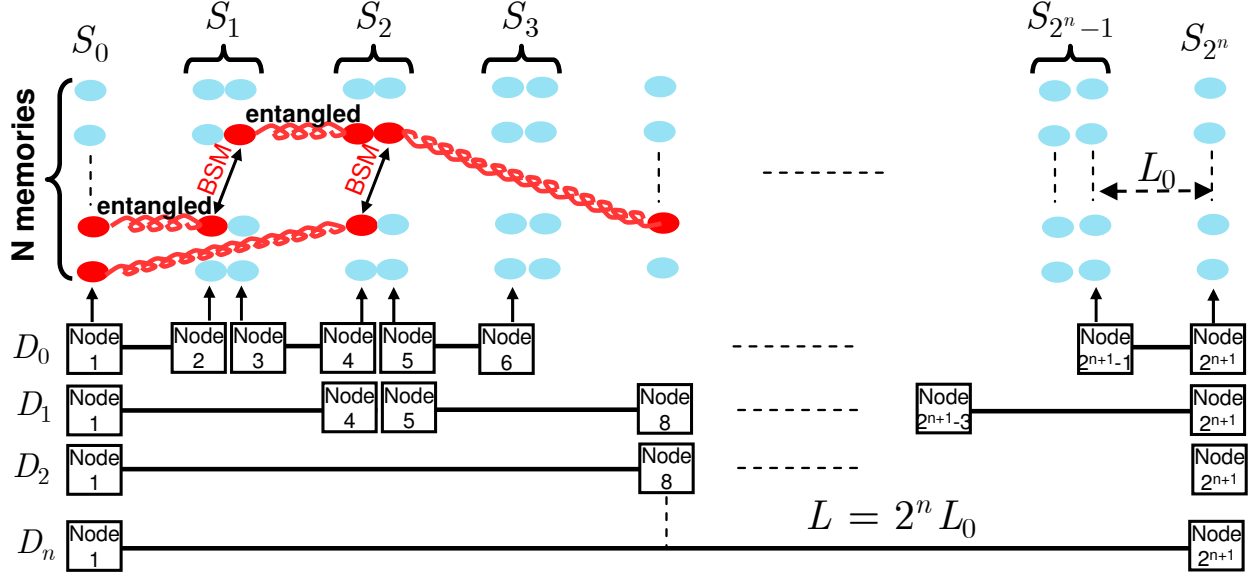
FIG. 1: (Color online) A quantum repeater with multiple quantum memories per node. At each round, we first employ an entanglement distribution protocol to entangle any unentangled memory pairs over D_0 links. It takes at least $T_{\rm ED} = L_0/c$ to learn about the success/failure of these attempts, which succeed with probability P_S, hence $T_{\rm ED}$ is the shortest period at which any protocol can run. At any such cycle, we also match up entangled pairs at different stations, according to our partial nesting protocol (see the text), to perform Bell-state measurements (BSMs), which succeed with probability P_M.

banks in Fig. 1 by nodes 1 to 2^{n+1}. Nodes $2i$ and $2i + 1$, for $i = 1, \ldots, 2^n - 1$, form physical stations S_i at which Bell-state measurements (BSMs) take place. For future reference, we also denote the family of sublinks of length $2^k L_0$, corresponding to nesting level $k = 0, \ldots, n$, that connect nodes $2^{k+1}(i - 1) + 1$ and $2^{k+1}i$, for $i = 1, \ldots, 2^{n-k}$, by D_k, and the set of all stations corresponding to nesting level $k > 0$ by $S^{(k)} \equiv \{S_{2^{k-1}(2l-1)}, l = 1, \ldots, 2^{n-k}\}$.

The probabilistic framework that we use for our rate analysis is as follows. We denote the probability of success for our employed entanglement distribution scheme over distance L_0 by P_S. The implicit assumption here is that our entanglement distribution scheme is heralding, i.e., there is a mechanism by which we can verify or become informed of whether our entangling attempt has failed or succeeded [2]. We also assume that the whole process of entanglement distribution from the time that it starts until we learn about its result takes $T_{\rm ED} \equiv L_0/c$. This is the fundamental time period at which we can attempt successively to

4

entangle two specific memories at distance L_0. We further assume that, in the absence of memory errors, the memories A and B, upon a successful entangling attempt, are in the maximally entangled state $|\psi^+\rangle_{AB}$, where $|\psi^\pm\rangle_{AB} = (|01\rangle_{AB} \pm |10\rangle_{AB})/\sqrt{2}$ and $|j\rangle_K$ is the logical qubit $j = 0, 1$ for memory $K = A, B$. Similarly, we associate a success probability $P_M < 1$ to each BSM. This probability, in principle, depends on the state of the memories before performing the BSM. In our optimistic analysis, we assume P_M is the maximum probability of obtaining a conclusive result for a given BSM module. For instance, if we rely on a specific pattern of photodetection events for performing a BSM, the inefficiency of photodetectors may result in inconclusive outcomes, in which case we consider the BSM has failed.

The probabilistic nature of our measurement modules imposes stringent conditions on how a quantum repeater protocol may run. In the nested purification protocol (NPP) proposed in [1], in order to extend the existing entanglement on D_{k-1} links to D_k links, we must perform BSMs at stations $S^{(k)}$, for $k = 1, \ldots, n$. Suppose that we have access to deterministic gates for local operations and measurements, and that no purification is required. Then, after establishing entanglement over all D_0 links, one can potentially perform all required BSMs simultaneously at all stations. In our probabilistic setup, however, there is no guarantee that the BSM at one nesting level provides us with entangled states required for the next nesting level. More generally, in a multiple-memory configuration, it is important to know on which pair of memories a BSM can be performed, and for that one needs to communicate between the nodes. This is also the case for the NPP if the employed purification scheme is probabilistic [11].

The above observation requires us to distinguish between two types of BSMs that one can perform. In our protocol, we consider a BSM at nesting level $k = 1, \ldots, n$ to be "informed" if it is performed on a pair of memories in nodes $2i$ and $2i + 1$ *known* to station $S_i \in S^{(k)}$ to be entangled with one of the memories in the relevant distant locations. A BSM is called "blind" if the above condition does not necessarily hold [7].

With the above considerations, our mth-order partial nesting protocol (m-PNP), for $m = 1, \ldots, n$, proceeds as follows. The m-PNP is a cyclic protocol with period T_{ED}. At each T_{ED}-long cycle, we first attempt to entangle unused memories of D_0 links. We learn about the success/failure of these attempts by the next cycle. At each cycle, intermediate stations also match up as many memories as possible in their two banks for performing
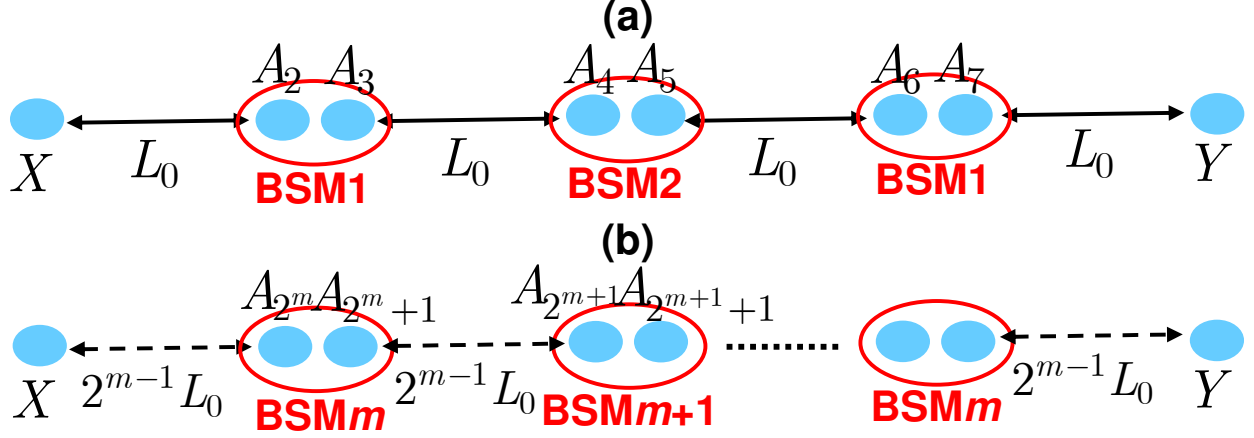
5

FIG. 2: (Color online) (a) A quantum repeater link with nesting level $n = 2$. Here, we have only shown memories $A_2, \ldots, A_7$ that contribute to entanglement generation between memories X and Y. In the 2-PNP, all measurements are informed so if measurements labeled by BSM1 occur at time t_1, BSM2 will occur at time $t_1 + T_1$. In the 1-PNP, however, all measurements happen at time t_1. In such a case, BSM2 will be a blind BSM. (b) A snapshot of memories involved in creating entanglement between X and Y in the general case of m-PNP. The assumption here is that we know that X and A_{2^m}, A_{2^m+1} and $A_{2^{m+1}}$, and so on are entangled because all previous measurements have been informed. The remaining measurements in $2^n/2^{m-1} - 1$ stations shown above all occur at the same time. Among these BSMs, those measurements that correspond to nesting level m, labeled by BSMm are informed, and the rest are blind.

BSMs. In our protocol, we perform *informed* BSMs at stations $S^{(k)}$, for $k = 1, \ldots, m$, on memories known to be entangled over D_{k-1}, and perform *blind* BSMs on memories known to be entangled over D_{m-1}, but not necessarily over D_{k-1}, at stations $S^{(k)}$, for $k = m+1, \ldots, n$; see Fig. 2. These blind BSMs are performed as soon as such measurements can be done. Purification may or may not be used over the first m nesting levels. After each measurement, or after creating entanglement over distance L, the involved memories will be released to be used again in the process of entanglement distribution. The required time for quantum measurement/processing is considered to be negligible throughout the paper.

The key to the polynomial scaling of the rate in quantum repeaters is in performing *informed*, rather than blind, BSMs. That can better be understood by comparing the special case of $m = n$, which corresponds to the probabilistic NPP, with quantum relays [16]. In quantum relays, one attempts to establish entanglement over D_0 and subsequently performs

BSMs at all stations without learning about the success/failure of previous entangling attempts. Quantum relays can, in principle, be run at a rate faster than $1/T_{\text{ED}}$ and may not need to use quantum memories, but their overall rate of entanglement generation will be proportional to $P_S^{2^n} P_M^{2^n-1} \propto (P_S P_M)^{L/L_0}$, which, for $P_S P_M < 1$, is exponentially decaying with L. This is in contrast with the probabilistic NPP, in which, as we will show, for ideal quantum memories, the rate scales polynomially with L as $P_S P_M^n \propto (L/L_0)^{\log_2 P_M}$.

Informed BSMs, however, require us to wait for classical signals. Even if we neglect the operation time for local gates/measurements, as we do throughout the paper, it still takes $T_k \equiv 2^{k-1} T_{\text{ED}}$, for $k = 1, \dots, n$, to transfer information from a station in $S^{(k)}$ to its closest station in $S^{(k+1)}$. This implies that there is a minimum delay of $T_{\text{ED}} + \sum_{k=1}^{m-1} T_k = T_m$, in the m-PNP, between our initial entangling attempts and the time that we can perform informed BSMs at the mth nesting level, during which memory decay is in effect. The memory decay will effectively reduce the rate of entanglement generation. An optimum rate may then be obtained by a combination of informed BSMs in the first few nesting levels and blind BSMs in the remaining final nesting levels.

To minimize the delay incurred by classical communication and the probabilistic nature of entanglement distribution/connection, we have shown that we must use a large number of memories per each node of our quantum repeater [9, 10]. To find an optimistic estimate of the rate, we then assume that $N \gg 1$; the rate behavior for a finite number of memories has been addressed in [8, 9, 10].

III.   RATE ANALYSIS

In this section we obtain optimistic estimates for the normalized rate, the rate of entanglement generation per employed memory, in the two cases of perfect and imperfect memories. In the latter case, the only source of error considered is memory dephasing, whose effect on the rate is studied in conjunction with different types of purification.

A.   Ideal memories

Let us first calculate the rate in the case of ideal quantum memories, i.e., when there is no source of error, and, therefore, no need for purification in the system. In the m-PNP,

7

starting with no entanglement at time zero, and for $N \gg 1$, the average number of entangled pairs over D_n that we create at time T_m, is given by $N_m^{(n)} \equiv NP_S P_M^{(2^{n-m+1}+m-2)}$. Here, NP_S is the average number of entangled pairs over each D_0 link at time T_{ED}, $NP_S P_M^{m-1}$ is the average number of entangled pairs over each D_{m-1} link after performing $m-1$ levels of informed BSMs, and $P_M^{(2^{n-m+1}-1)}$ accounts for the remaining BSMs, where 2^{n-m} of which are informed BSMs for the mth nesting level and the rest are blind; see Fig. 2(b). At time T_{ED}, there are on average $N - NP_S$ unused memory pairs in each D_0 link, which can be used to create new entangled pairs. We overestimate $N - NP_S$ by N, and repeat the above argument to obtain the same value $N_m^{(n)}$ for the average number of entangled pairs created over distance L at time $T_{\mathrm{ED}} + T_m$. By reusing the same argument for each cycle, in the steady state, $N_m^{(n)}$ gives an optimistic estimate of the average number of entangled pairs created per cycle. That implies that the steady-state rate of entanglement generation per each of $N2^{n+1}$ *ideal* memories used in the m-PNP, $m = 1, \cdots, n$, is approximated by

$$Q_m^{(n)} \equiv \frac{N_m^{(n)}}{T_{\mathrm{ED}} N 2^{n+1}} = \frac{P_S P_M^{(2^{n-m+1}+m-2)}}{2L/c}. \tag{1}$$

Here, $m = n$ represents the special case of the probabilistic NPP, which is the optimum scenario when memories are perfect.

B. Imperfect memories

Now, let us consider the case of imperfect memories. Here, we model the memory degradation by a dephasing process in which a qubit state $\hat{\rho}_A$, of memory A, is mapped, after decaying for a time period t, to

$$\Gamma_t^A(\hat{\rho}_A) = p(t/2)\hat{\rho}_A + [1 - p(t/2)]\hat{Z}_A \hat{\rho}_A \hat{Z}_A, \tag{2}$$

where $\hat{Z}_A$ is the Pauli Z-operator acting on memory A, $p(t) = [1 + \exp(-t/\tau_c)]/2$, and τ_c is the memory coherence time. The above dephasing process maps $\hat{\pi}_{AB}^{\pm} \equiv |\psi^{\pm}\rangle_{AB}\langle\psi^{\pm}|$ to the following rank-two Bell-diagonal state

$$\Gamma_t^A \otimes \Gamma_t^B(\hat{\pi}_{AB}^{\pm}) = p(t)\hat{\pi}_{AB}^{\pm} + [1 - p(t)]\hat{\pi}_{AB}^{\mp} \equiv \hat{\rho}_{AB}^{\pm}(t), \tag{3}$$

with fidelity $p(t)$. We can also show that a BSM on memories B and C of a four-memory system initially in the state $\hat{\rho}_{AB}^{+}(t) \otimes \hat{\rho}_{CD}^{+}(t)$ leaves A and D, up to a local unitary, in the state $\hat{\rho}_{AD}^{+}(2t)$.

Using the above model, we estimate $R_m^{(n)}$, the generation rate of maximally entangled states, per employed memory, in the presence of dephasing errors. We first consider the no-purification case, and then allow to use purification, without restricting ourselves to any specific purification protocol. In each case, for any memories X and Y at distance L in the m-PNP, we obtain the final state $\hat{\rho}_{f,XY}$ upon establishment of entanglement between them. We denote by A_i the memory in node $i = 2, \ldots, 2^{n+1} - 1$ that contributes to the successful creation of entanglement between X and Y; see Fig. 2.

To underestimate the decay effect, we make a further assumption in calculating $\hat{\rho}_{f,XY}$. In the m-PNP, after performing the last set of BSMs, corresponding to nesting levels m and higher, we must in principle wait, before claiming that X and Y are entangled, until we receive the classical data from all middle nodes. The effective state in such a case must then include the decay during this last step of communication. In certain applications such as QKD, however, we can perform the required QKD measurements on X and Y concurrent with the time that blind BSMs are performed, and later confirm whether they are entangled or not. Under the latter assumption, we do not need to account for the classical delays that occur after the last set of BSMs corresponding to nesting levels m and higher. Throughout the rest of this Section, we assume that we are operating under such conditions.

With the above assumptions, we overestimate the number of maximally entangled states that can be obtained out of $N_m^{(n)}$ copies of $\hat{\rho}_{f,XY}$ by $N_m^{(n)} E_C(\hat{\rho}_{f,XY})$. For a state $\hat{\rho}$, $E_C(\hat{\rho})$ is the entanglement cost, viz. the ratio M/K for starting with $(\hat{\pi}^+)^{\otimes M}$ and obtaining $\hat{\rho}^{\otimes K}$, in the limit of large K, using local operations and classical communication (LOCC). It is known that for the asymptotic version $E^\infty(\hat{\rho})$ of any "well-behaved" entanglement measure, we have $E_D(\hat{\rho}) \leq E^\infty(\hat{\rho}) \leq E_C(\hat{\rho})$, where $E_D(\hat{\rho})$ is the distillable entanglement, i.e., the ratio M/K for distilling $\hat{\rho}^{\otimes K}$ to $(\hat{\pi}^+)^{\otimes M}$, in the limit of large K, using LOCC [14]. Hence, our choice of entanglement cost as our entanglement measure is in line with our other optimistic assumptions. For $\hat{\rho}_{XY}^+(t)$, we have $E_D(\hat{\rho}_{XY}^+(t)) = 1 - H(p(t))$ and $E_C(\hat{\rho}_{XY}^+(t)) = H(\frac{1}{2} + \sqrt{p(t)[1 - p(t)]})$, where $H(p) = -p \log_2 p - (1 - p) \log_2(1 - p)$ [14]. Here, $E_D(t)$ also gives the asymptotic yield for the one-way hashing protocol for purifying $\hat{\rho}_{XY}^+(t)$ [12].

1. Without Purification

Let us first consider the case where no purification is employed. Suppose that we attempt to entangle D_0 links at time zero. Then by the time that we become informed of the establishment of entanglement, memories have already decayed for a period T_{ED}. For instance, X and A_2 are in the state $\hat{\rho}^+_{XA_2}(T_{ED})$. At $T_{ED} = T_1$, we also perform the BSMs corresponding to the first nesting level. That leaves X and A_4, right after performing the BSM and up to a local unitary, in $\hat{\rho}^+_{XA_4}(2T_1)$. It takes T_1 to inform A_4 of the BSM success, hence at time $T_2 = T_{ED} + T_1$, X and A_4 are in $\hat{\rho}^+_{XA_4}(3T_1)$. By continuing the same argument, we can show that at time T_m, $\hat{\rho}_{f,XY} = \hat{\rho}^+_{XY}(t_m^{(n)})$, where $t_m^{(n)} \equiv T_{n+1} + (m-1)T_n$, for $m = 1, \dots, n$. By time T_m, we do not necessarily know whether X and Y are entangled, but as mentioned before, in some applications such as QKD, we can often use X and Y at time T_m, and later verify whether they are entangled or not. Under this assumption, our optimistic rate is given by

$$R_m^{(n)} \approx Q_m^{(n)} E_C(\hat{\rho}^+_{XY}(t_m^{(n)})), \quad \text{(no purification)}. \tag{4}$$

Given that, to the first order, $E_C(\hat{\rho}^+_{XY}(t)) \propto t \exp(-2t/\tau_c)$ for $t \gg \tau_c$, and that $t_m^{(n)} \geq L/c$, the rate in Eq. (4) exponentially decays with distance for a finite value of τ_c. This implies that the partial nesting protocol cannot help us avoid the exponential decay of the normalized rate with distance if we do not use any purification.

2. With Purification

Now, let us consider the case when we allow to use purification. In the m-PNP, our last chance to purify memories is in the mth nesting level. In order to obtain an upper bound on the rate, suppose that at the nesting level $m - 1 \geq 1$, we are provided with maximally entangled states $\hat{\pi}^+$ over D_{m-2} links. (We consider the special case of $m = 1$ later.) The BSMs at stations $S^{(m-1)}$, upon success, will leave us, up to a local unitary, in $\hat{\pi}^+$. There will be, however, a delay of T_{m-1} to inform stations $S^{(m)}$ of the success/failure of the above BSMs, during which the memories decay to $\hat{\rho}^+(T_{m-1})$. In the special case of $m = 1$, the minimum delay is given by T_{ED}, hence the decayed state will be $\hat{\rho}^+(T_{ED})$. The rate is then optimistically approximated by

$$R_m^{(n)} \approx Q_m^{(n)} E_C(\hat{\rho}^+_{XY}(T_m^{(n)})), \quad \text{(with purification)}, \tag{5}$$

where $T_m^{(n)} = \max(T_{\mathrm{ED}}, T_{m-1})$. Here, for a fixed value of n, $E_C(\hat{\rho}_{XY}^+(T_m^{(n)}))$ is a decreasing function of m, whereas $Q_m^{(n)}$ is an increasing function of m. That implies that, for fixed values of τ_c and L, there are optimum values of m and n that maximize the normalized rate, and such a maximum may scale better than exponentially with distance, as we show next.

In order to find the asymptotic behavior of $R_{\mathrm{opt}} \equiv \max_{m,n} R_m^{(n)}$ in the limit of large distances, we consider all possible choices that we have for our system parameters m and n. The only additional assumption in our analysis is that $P_S \propto \exp(-\alpha L_0)$, where α is inversely proportional to the attenuation length of the channel. This is a reasonable assumption as most proposed entanglement distribution schemes require the transmission of a single photon along the channel [2, 5, 6]. Now let us consider the following cases for the optimum values of m and n, respectively, denoted by m_{opt} and n_{opt}, all in the limit of large distances:

(1) Suppose n_{opt} is finite, then, because $P_S \propto \exp(-\alpha L/2^n)$, R_{opt} degrades exponentially with L.

(2) Suppose $n_{\mathrm{opt}} \to \infty$ but $n_{\mathrm{opt}} - m_{\mathrm{opt}}$ is finite. Then the entanglement cost term in Eq. (5), will scale as $\exp[-L/(c\tau_c 2^{n_{\mathrm{opt}} - m_{\mathrm{opt}} + 1})]$, which again represents an exponential decay of the normalized rate with distance.

(3) Suppose $n_{\mathrm{opt}} \to \infty$ and $n_{\mathrm{opt}} - m_{\mathrm{opt}} \to \infty$, then

$$R_{\mathrm{opt}} \propto \exp(-\alpha L_0^{\mathrm{opt}}) \exp\left(\frac{L \ln(P_M)}{2^{m_{\mathrm{opt}} - 1} L_0^{\mathrm{opt}}}\right) E_C(\hat{\rho}_{XY}^+(2^{m_{\mathrm{opt}} - 2} L_0^{\mathrm{opt}}/(c\tau_c))), \qquad (6)$$

where $L_0^{\mathrm{opt}} = L/2^{n_{\mathrm{opt}}}$. (Note that $R_1^{(n)} \le R_2^{(n)}$, therefore $T_{m_{\mathrm{opt}}}^{(n_{\mathrm{opt}})} = 2^{m_{\mathrm{opt}} - 2} L_0^{\mathrm{opt}}/c$.) Now, if $2^{m_{\mathrm{opt}}} L_0^{\mathrm{opt}}$ is finite, then the entanglement cost term in the above equation will be a constant, whereas the second term, so long as $P_M < 1$, decays exponentially with L. If $2^{m_{\mathrm{opt}}} L_0^{\mathrm{opt}} \to \infty$, we can then replace the entanglement cost term in Eq. (6) with its asymptotic value to obtain

$$R_{\mathrm{opt}} \propto \exp(-\alpha L_0^{\mathrm{opt}}) \exp\left(\frac{L \ln(P_M)}{2^{m_{\mathrm{opt}} - 1} L_0^{\mathrm{opt}}}\right) \exp(-2^{m_{\mathrm{opt}} - 1} L_0^{\mathrm{opt}}/(c\tau_c)). \qquad (7)$$

In the above equation, there are two competing terms: the second term that decays exponentially with $L/(2^{m_{\mathrm{opt}} - 1} L_0^{\mathrm{opt}})$ and the third term that decays exponentially with $2^{m_{\mathrm{opt}} - 1} L_0^{\mathrm{opt}}$. The best rate-versus-distance scaling can be achieved when these two terms scale the same with L. That can be achieved by assuming $2^{m_{\mathrm{opt}} - 1} L_0^{\mathrm{opt}} = \sqrt{\beta L}$, for some $\beta > 0$, in which case both above-mentioned terms decay exponentially with $\sqrt{L}$. After optimizing over β

and assuming that $m_{\text{opt}} \to \infty$, we find that

$$R_{\text{opt}} \propto \exp\left(-2\sqrt{\frac{L\ln(1/P_M)}{c\tau_c}}\right), \tag{8}$$

which is exponentially decaying with $\sqrt{L}$, rather than L as obtained for cases (1) and (2). This is in fact the best rate-versus-distance scaling achievable for our quantum repeater system under our optimistic assumptions. Given that the above optimized rate is obtained in the regime of $m_{\text{opt}}, n_{\text{opt}} \to \infty$, we can further manipulate the ratios $R_{m+1}^{(n+1)}/R_m^{(n)}$ and $R_{m+1}^{(n)}/R_m^{(n)}$ to find that, for large distances, the optimum rate is obtained at

$$L_0^{\text{opt}} \approx 2\ln P_M^{-1}/\alpha, \tag{9}$$

which is a constant identical to what one obtains for the ideal-memory case, and

$$2^{m_{\text{opt}}} \approx \alpha\sqrt{2Lc\tau_c/\ln P_M^{-1}}. \tag{10}$$

This shows that the optimum rate is obtained when $2^{n_{\text{opt}}} \propto L$ and $2^{m_{\text{opt}}} \propto \sqrt{L}$. Such an improvement over the no-purification case is associated with the employed purification scheme as we discuss next.

The proposed purification schemes can be divided into one-way and two-way protocols [12]. In most proposed one-way protocols, it is commonly the case that after performing proper local operations on one side, the measurement results are sent to the other side at which, after further manipulation, the purified states are obtained. In other words, the operation of such schemes is equivalent to that of a quantum code, in which errors occurred during both the processing as well as the storage time may be corrected. In most proposed two-way purification protocols, it is commonly the case that we perform local operations on both sides, and exchange the measurement results between the two parties. The purification process can then be completed only after the receipt of classical data. The problem with these protocols is that the faulty memories will decay during the transmission of classical data. That may result in an exponential decay of rate with distance as detailed below.

To see the effect of a typical two-way purification scheme on the normalized rate, suppose that in order to purify $\hat{\rho}_{AB}$, an arbitrary two-qubit state between two partners A and B at distance l, we apply local operations/measurements on M copies of $\hat{\rho}_{AB}$ to obtain $K < M$ copies of $\hat{\rho}_{AB}^{\text{pur}}$. Then, we communicate between A and B, to verify whether the purification is

successful. Upon success we perform another set of local operations, modeled by a quantum operation Λ, on $(\Gamma^A_{l/c} \otimes \Gamma^B_{l/c})^{\otimes K}[(\hat{\rho}^{\text{pur}}_{AB})^{\otimes K}]$. For many schemes proposed in [11], Λ is just the identity operator due to a discard-or-keep action, so we further assume that the dephasing process modeled by $\Gamma^A_{l/c} \otimes \Gamma^B_{l/c}$ commutes with Λ. Under these conditions, we show that the fidelity of memories A and B after purification cannot exceed $p(l/c)$. Suppose $\hat{\rho}^F_{AB} \equiv \Lambda(\hat{\rho}^{\text{pur}}_{AB})$, where $F = \text{tr}(\hat{\pi}^+ \hat{\rho}^F_{AB})$ is the fidelity of the state after purification without considering the memory decay. The fidelity of such state after memory decay for a period $t = l/c$ will then satisfy

$$
\begin{aligned}
\text{tr}[\hat{\pi}^+_{AB}\Gamma^A_t \otimes \Gamma^B_t(\hat{\rho}^F_{AB})] &= \text{tr}[\Gamma^A_t \otimes \Gamma^B_t(\hat{\pi}^+_{AB})\hat{\rho}^F_{AB}] \\
&= \text{tr}[(p(t)\hat{\pi}^+_{AB} + [1-p(t)]\hat{\pi}^-_{AB})\hat{\rho}^F_{AB}] \\
&\leq p(t)F + [1-p(t)](1-F) \leq p(t), \quad\quad (11)
\end{aligned}
$$

where the first equality comes from the cycling property of the trace operator, and equality holds for the first inequality if $\hat{\rho}^F_{AB} = F\hat{\pi}^+_{AB} + (1-F)\hat{\pi}^-_{AB}$. We used $F \leq 1$ in the last inequality. This implies that even if $\hat{\rho}^{\text{pur}}_{AB} = \hat{\pi}^+_{AB}$, by the end of the above purification procedure, we ideally end up with $\hat{\rho}_A B^+(l/c)$. Using such purification schemes in the m-PNP will then leave memories X and A_{2^m}, at best, in the state $\hat{\rho}^+_{X,A_{2^m}}(T_m)$, right before performing the last set of BSMs. This is the case too for other similar pairs of memories. After performing the remaining BSMs, we end up with $\hat{\rho}_{f,XY} = \hat{\rho}^+_{XY}(T_{n+1})$, for which the entanglement cost scales as $\exp[-2L/(c\tau_c)]$, and that proves our claim.

Whereas the two-way purification schemes considered above cannot provide us with the optimum rate-versus-distance scaling shown in Eq. (8), there are one-way purification schemes that achieve a similar scaling. For instance, in the m-PNP, suppose we perform no purification up until the m-th nesting level, at which point we perform a one-way hashing protocol [12] on entangled states distributed over D_{m-1}. These entangled memories, before applying the purification and up to a local unitary, are in the state $\hat{\rho}^+(t^{(m)}_m/2)$. To perform the hashing protocol, we need to perform encoding measurements on one side of each entangled link over D_{m-1} and send the results to the nodes on the other side of the link. Suppose the nodes that receive this encoding information are those at which the mth nesting level measurements will occur. For all other nodes, we not only can perform, at time T_m, the measurements required for the hashing protocol but also, at the same time, we can perform those blind BSMs, according to our partial nesting protocol, needed for entanglement swapping as

well as the required measurements on the end nodes with regard to our QKD application. By this trick, only memories located at $S^{(m)}$ will undergo an additional T_m-long decay due to the transmission delay before getting purified. The normalized rate after performing the above purification as well as the remaining BSMs is then given by $Q_m^{(n)} E_D(\hat{\rho}_{XY}^+(t_m^{(m)}/2 + T_m/2))$, which, for $m = 1$ and $L_0 \propto \sqrt{L}$, degrades exponentially with $\sqrt{L}$ similar to Eq. (8). Here, we used the fact the asymptotic yield of the one-way hashing protocol for our decayed state is given by the entanglement of distillation $E_D(\hat{\rho}_{XY}^+(t)) \propto \exp(-2t/\tau_c)$, for $t \gg \tau_c$.

The above example shows that with even one step of purification we can achieve the optimum scaling achievable by a quantum repeater system that only relies on entanglement purification for mitigating the effect of memory errors. This is a weaker requirement than what needed for fully fault-tolerant schemes, which rely on the frequent use of error correction schemes. Although here we assumed that the gates used in the purification scheme are deterministic and error-free, we can possibly relax this assumption by using probabilistic gates for purification. How the rate decays in this new scenario is a matter of further investigation.

IV. NUMERICAL RESULTS

In this section, we numerically study the effect of different system parameters on the normalized rate. In all graphs presented, we assume $P_S = 0.2 \times 10^{-0.01 L_0}$, which is an appropriate choice for the schemes proposed in [2, 5]. Furthermore, The underlying physical channel is assumed to be an optical fiber with $0.17\,\mathrm{dB/km}$ loss, for which $c = 2 \times 10^8\,\mathrm{m/s}$ and $1/\alpha = 50\,\mathrm{km}$. For numerical purposes, we have used the normalized rate given by Eqs. (4) and (5), for, respectively, systems without and with purification.

Figure 3(a) shows the rate in Eq. (5) versus nesting level n for $L = 1000\,\mathrm{km}$, hence $L/c = 5\,\mathrm{ms}$. For each value of n, we have optimized m to maximize the rate. The optimum values of m turned out to be equal to n for all the marked points in Fig. 3(a). This figure clearly shows the criticality of L/c as a benchmark for the required coherence time. Whereas the rate drops over an order of magnitude when the coherence time is reduced from $5\,\mathrm{ms}$ to $1\,\mathrm{ms}$, there is only a slight improvement in the rate if we use memories with $100\,\mathrm{ms}$ coherence time. We should bear in mind, however, that these values are only valid in the case of $N \gg 1$. The benefit of using quantum memories with longer coherence time is that
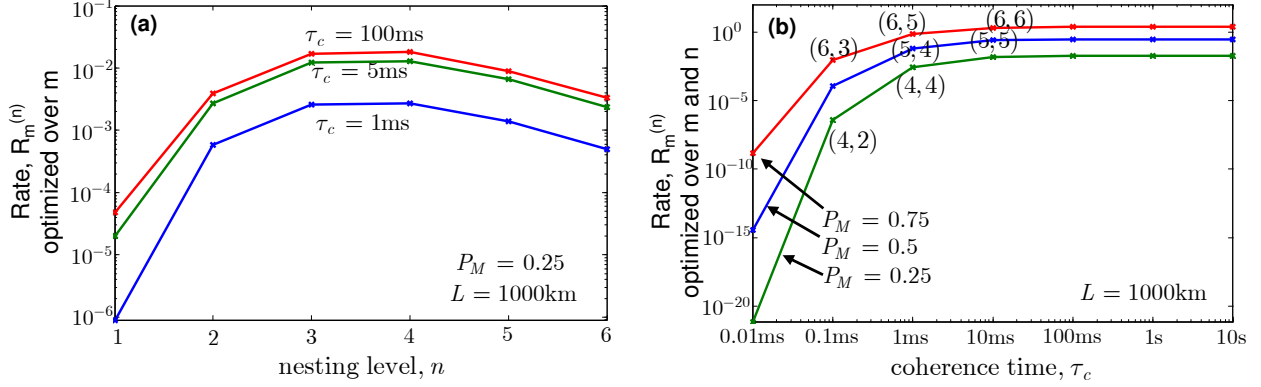
FIG. 3: (Color online) An optimistic estimate to the generation rate of maximally entangled states per employed memory, $R_m^{(n)}$, for the m-th order partial nesting protocol, in the presence of memory dephasing. The implicit assumption is that we employ a sufficiently large number of memories to minimize the waiting times due to classical communication. In (a), we have plotted $R_m^{(n)}$, optimized over m, versus the nesting level, for three different values of coherence time. In (b), we have plotted $R_m^{(n)}$ versus coherence time, optimized over m and n. For fixed values of τ_c, the rate scales exponentially in $\sqrt{L}$. The pairs (n, m), on several points on the graph, show the optimum nesting levels, n, and the optimum numbers of informed BSMs, m, at those points. In both (a) and (b), we assume that our entanglement distribution succeeds with probability $P_S = 0.2 \times 10^{-0.01 L_0}$ and the total distance is $L = 1000\,\mathrm{km}$, thus $L/c = 5\,\mathrm{ms}$ in optical fibers.

we can still achieve the same value of $R_m^{(n)}$ with a lower number of employed memories [10]. Another interesting point in Fig. 3(a) is that the achievable rate for both values of $n = 3$ and $n = 4$ is about the same. Although, numerically, $n = 4$ attains the maximum rate, for practical purposes, it is beneficial to use the less costly architecture corresponding to $n = 3$. One last point about this figure is the difference in the slope of rate for values of n smaller than the optimum value of n, and those that are larger. It is because for $n < n_{\mathrm{opt}}$, loss is the dominant factor, whereas for $n > n_{\mathrm{opt}}$ the BSM success probability brings the rate down.

Figure 3(b) shows the optimum rate versus coherence time for three values of P_M at $L = 1000\,\mathrm{km}$. On several points on the graph, we have indicated pairs (n, m) that achieve the maximum achievable rate. For $\tau_c \ll L/c$, the partial nesting protocol favors using blind BSMs as they require memories with shorter coherence times. For lower values of P_M, one needs to use fewer nesting levels to attain the maximum achievable rate. The optimum values of n in the above cases agree with the relation $L_0^{\mathrm{opt}} \approx 2\ln(P_M^{-1})/\alpha$ found in the
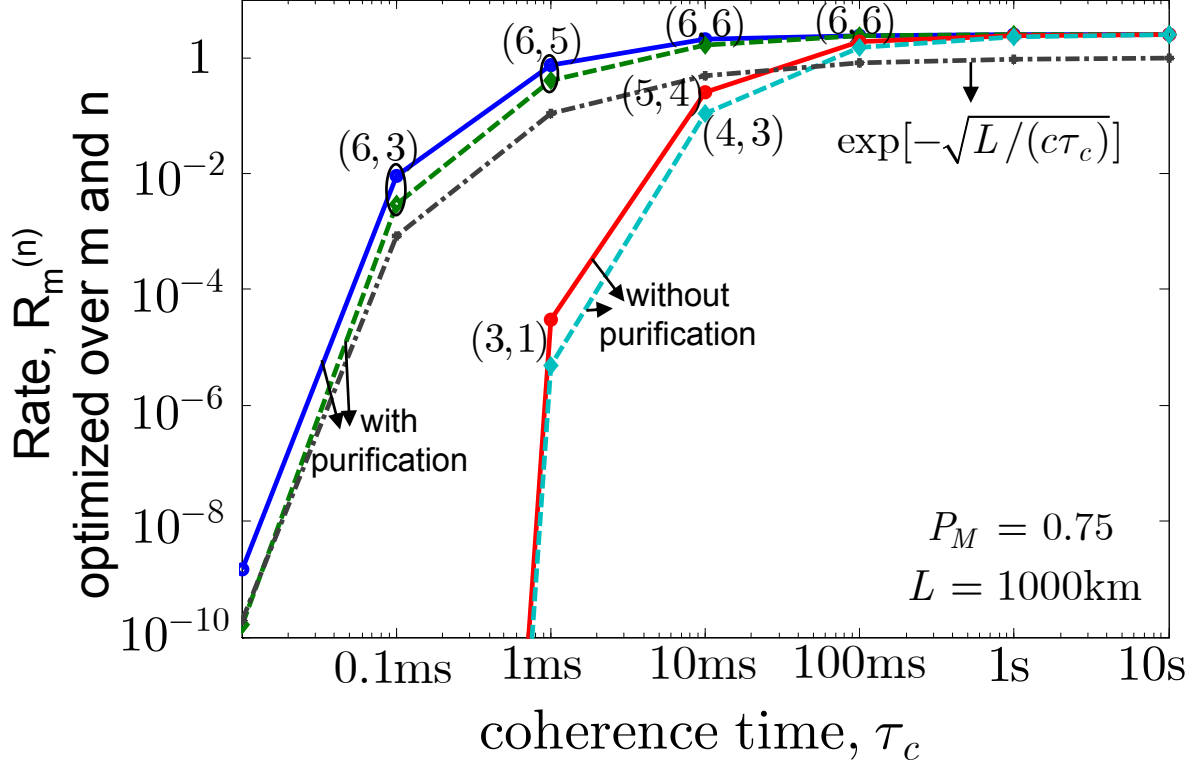
FIG. 4: (Color online) Estimates for the generation rate of maximally entangled states per employed memory, $R_m^{(n)}$, versus coherence time, optimized over m and n, for $P_M = 0.75$, $P_S = 0.2 \times 10^{-0.01 L_0}$, $L = 1000 \,\mathrm{km}$, and $L/c = 5 \,\mathrm{ms}$. The employed entanglement measures in these graphs are entanglement cost for the solid lines, and entanglement of distillation for the dashed lines. For fixed values of τ_c, the curves scale exponentially in L if no purification is used and exponentially in $\sqrt{L}$ with purification; see dash-dotted line for comparison. The two-tuples (n, m), on several points on the graph, show the optimum nesting levels, n, and the optimum numbers of informed BSMs, m, at those points.

previous section.

Figure 4 demonstrates the rate behavior in both limits of long and short coherence times. The solid lines represent the optimistic normalized rate when entanglement cost is used as our entanglement measure as in Eqs. (4) and (5). The dashed lines represent the achievable normalized rate when entanglement of distillation is used instead. It can be seen that the difference between the two cases is small. It is noticeable that the normalized rate, regardless of the employed entanglement measure or purification scheme, approaches a constant value, $\max_n Q_n^{(n)}$, when $\tau_c \gg L/c$, in accord with what reported in [7, 8] and Eqs. (4) and (5). In

the other unexplored extreme of $L \gg c\tau_c$, however, the situation is much different. In the no-purification case, because, to the first order, $E_C(\hat{\rho}_{XY}^+(t)) \propto t\exp(-2t/\tau_c)$ for $t \gg \tau_c$, the rate in Eq. (4) exponentially decays with distance. Proper use of purification helps, but only a little. As shown by the dash-dotted line in Fig. 4, at the optimum values of m and n, $R_m^{(n)}$ in Eq. (5) scales as a power of $\exp[-\sqrt{(L/c)/\tau_c}]$. The same effect has been shown in Fig. 5, where we have plotted the optimized rate versus L for several values of coherence time. It can be seen that for $\tau_c = 100\,\mathrm{ms}$, the logarithm of rate scales linearly with $\log L$, which represent a polynomial scaling of rate with distance. For short coherence times, however, we will observe the exponential decay with $\sqrt{L}$ as discussed before. Another interesting point is the optimum values of m and n, which scale, respectively with $\log \sqrt{L}$ and $\log L$.

V. CONCLUSIONS

In this paper, we studied quantum repeaters that relied on probabilistic schemes for entanglement distribution and connection. We introduced a proper rate-over-cost figure of merit for such systems by looking at the ratio between the rate of entanglement generation—when the system resources are being fully employed to successively create entangled states—and the total number of employed memories. We used entanglement measures to quantify the generation rate of entanglement. We believe that to properly compare different quantum repeater setups, one must study such a normalized rate instead of merely considering the fidelity of generated entangled states at the end of only one round of the protocol. We studied the effect of memory dephasing under the assumption that the only mechanism employed to mitigate errors is purification. We further assumed that the number of memories employed per node is sufficiently large such that, by parallel entangling attempts, we can minimize the unavoidable communication delay between relevant nodes. The memory decay, in principle, can be dealt with using complex fault-tolerant schemes [18], which may not be available in the near future, and has not been considered in our work. We showed that, whereas in the ideal case of infinitely long coherence times the normalized rate scales polynomially with distance, with imperfect quantum memories the rate-over-cost asymptotically scales, at best, as a power of $\exp[-\sqrt{(L/c)/\tau_c}]$, with τ_c being the memories' coherence time. This rate behavior is nevertheless superior to the fully exponential decay of rate with distance in quantum relays, which do not use quantum memories, or in systems that do not employ
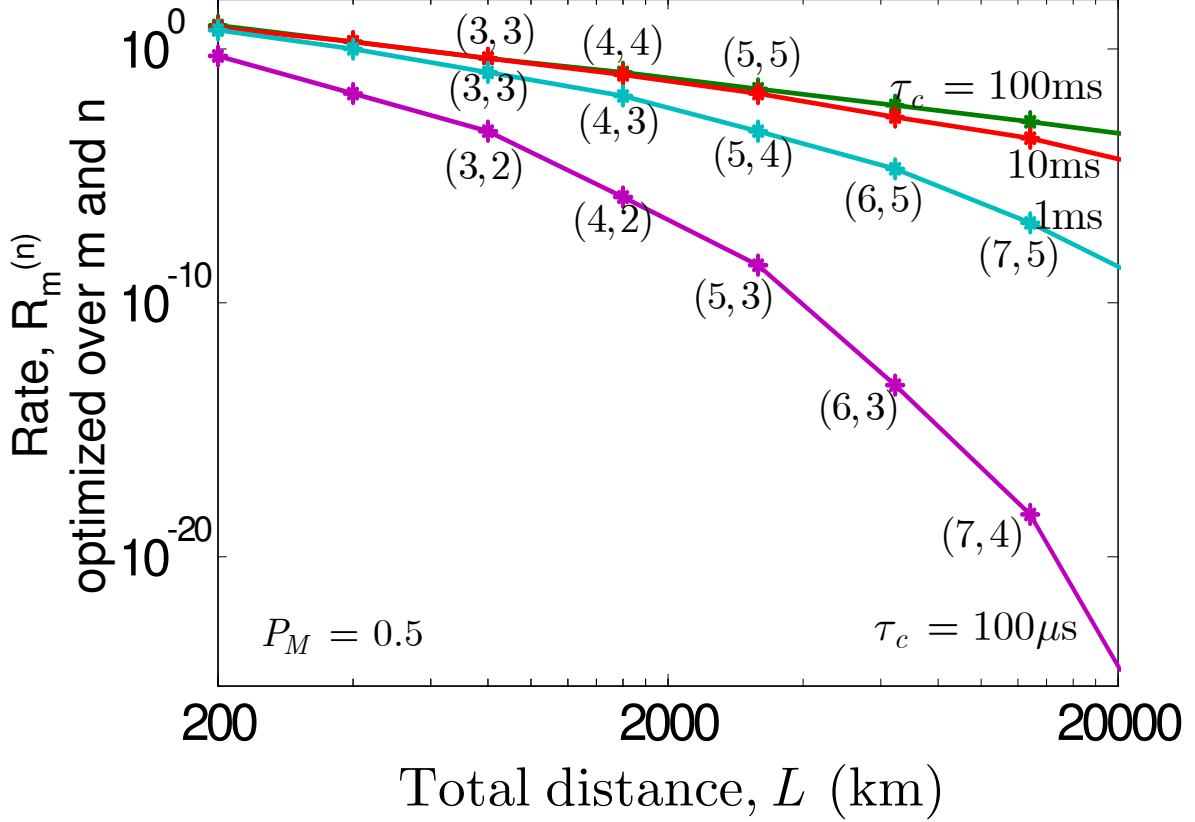
FIG. 5: (Color online) Estimates for the generation rate of maximally entangled states per employed memory, $R_m^{(n)}$, versus total distance L, optimized over m and n, for different values of coherence time at $P_M = 0.5$ and $P_S = 0.2 \times 10^{-0.01L_0}$. The employed entanglement measure in these graphs is the entanglement cost and a proper purification protocol is used. It can be seen that, for $\tau_c \gg L/c$, the curves scale polynomially with L, whereas, for $\tau_c \ll L/c$, they scale exponentially with $\sqrt{L}$. The two-tuples (n, m), on several points on the graph, show the optimum nesting levels, n, and the optimum numbers of informed BSMs, m, at those points. They, respectively, scale with $\log L$ and $\log \sqrt{L}$, for large distances.

purification but use decaying memories. Our analysis helps us estimate the minimal cost of our quantum communication systems. In the particular case of QKD, to create 1000 secure key bits/s over 1000 km, we may need thousands of memories with coherence times exceeding or on the order of 10 ms.

---

[1] H. J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **81**, 5932 (1998).

[2] L. M. Duan, M. D. Lukin, J. I. Cirac, and P. Zoller, Nature **414,** 413 (2001).

[3] M. Razavi and J. H. Shapiro, Phys. Rev. A **73,** 042303 (2006); J. Amirloo, M. Razavi, and A. H. Majedi, in Technical Digest, Int. Conf. on Quantum Inform., Paper JMB54, Boston, MA (2008).

[4] L. Childress *et al.*, Phys. Rev. Lett. **96,** 070504 (2006); T. D. Ladd *et al.*, New J. Phys. **8,** 134 (2006).

[5] N. Sangouard, C. Simon, J. Minàr, H. Zbinden, H. de Riedmatten, and N. Gisin, Phys. Rev. A **76,** 050301(R) (2007); C. Simon *et al.*, Phys. Rev. Lett. **98,** 190503 (2007).

[6] B. Zhao *et al.*, Phys. Rev. Lett. **98,** 240502 (2007); L. Jiang *et al.*, Phys. Rev. A **76,** 012301 (2007); N. Sangouard *et al.*, Phys. Rev. A **77,** 062301 (2008); M. Gao *et al.*, Phys. Rev. A **79,** 042301 (2009).

[7] L. Hartmann, B. Kraus, H.-J. Briegel, and W. Dür, Phys. Rev. A **75**, 032310 (2007).

[8] O. A. Collins, S. D. Jenkins, A. Kuzmich, and T. A. B. Kennedy, Phys. Rev. Lett. **98**, 060502 (2007).

[9] M. Razavi, H. Farmanbar, and N. Lütkenhaus, in *Technical Digest, Paper JWA48, OFC'08, San Diego, CA* (2008).

[10] M. Razavi *et al.*, Proc. SPIE **7236**-03, San Jose, CA (2009).

[11] C. H. Bennett *et al.*, Phys. Rev. Lett. **76**, 722 (1996); D. Deutsch *et al.*, Phys. Rev. Lett. **77**, 2818 (1996).

[12] C. H. Bennett *et al*, Phys. Rev. A **54,** 3824 (1996).

[13] W. Dür *et al.*, Phys. Rev. Lett. **91,** 107903 (2003).

[14] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **84**, 2014 (2000); G. Vidal, W. Dür, and J. I. Cirac, Phys. Rev. Lett. **89**, 027901 (2002).

[15] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[16] H. de Riedmatten *et al.*, Phys. Rev. Lett. **92**, 047904 (2004).

[17] B. Zhao *et al.*, Nature Physics **5,** 95 (2009); R. Zhao *et al.*, Nature Physics **5,** 100 (2009).

[18] L. Jiang *et al.*, Phys. Rev. A **79,** 032325 (2009).